Subscribe (Full Service) Register (Limited Service, Free) Login

Search: ● The ACM Digital Library ○ The Guide

USPTO

THE ACM DIGITAL LIBRARY

Feedback

searching similar documents jaccard distance
Terms used: searching similar documents jaccard distance

Sort results by relevance

Display results expanded form

Save results to a Binder

□ Open results in a new window

Refine these results wi
Try this search in The

Results 1 - 20 of 52 Result page: 1 2 3 next >>

1 The LIVE-project: retrieval experiments based on evaluation viewpoints
P. Bollmann, F. Jochum, U. Reiner, V. Weissmann, H. Zuse
June 1985 SIGIR '85: Proceedings of the 8th annual international ACM SIGIR conference on Research and development in information retrieval
Publisher: ACM
Full text available: pdf(150.62 KB) Additional Information: full citation, abstract, references, cited by

Bibliometrics: Downloads (6 Weeks): 2, Downloads (12 Months): 16, Citation Count: 12

Besides the operators 'and', 'or' and 'not' the GRIPS retrieval language contains thesaurus — operators to extend the query and truncation — and context-operators for freetext and Boolean searching. In a similar way several other viewpoints ...

2 Unified framework for fast exact and approximate search in dissimilarity spaces
Tomáš Skopal
November 2007 ACM Transactions on Database Systems (TODS), Volume 32 Issue 4
Publisher: ACM
Full text available: pdf(1.74 MB) Additional Information: full citation, abstract, references, index terms

Bibliometrics: Downloads (6 Weeks): 12, Downloads (12 Months): 253, Citation Count: 0

In multimedia systems we usually need to retrieve database (DB) objects based on their similar to a query object, while the similarity assessment is provided by a measure which defines a (dis similarity score for every pair of DB objects. In most ...

Keywords: Similarity retrieval, approximate and exact search

3 Query clustering using user logs
January 2002 ACM Transactions on Information Systems (TOIS), Volume 20 Issue 1
Publisher: ACM
Full text available: pdf(1.31 MB) Additional Information: full citation, abstract, references, cited by, index term: review

Bibliometrics: Downloads (6 Weeks): 22, Downloads (12 Months): 177, Citation Count: 34

Query clustering is a process used to discover frequently asked questions or most popular topic: a search engine. This process is crucial for search engines based on question-answering. Becaus of the short lengths of queries, approaches based on ...

Keywords: Query clustering, search engine, user log, web data mining

4 Temporal profiles of queries
Rosie Jones, Fernando Diaz

July 2007 ACM Transactions on Information Systems (TOIS), Volume 25 Issue 3
Publisher: ACM
Full text available: pdf(430.31 KB) Additional Information: full citation, abstract, references, index terms

Bibliometrics: Downloads (6 Weeks): 12, Downloads (12 Months): 316, Citation Count: 1

Documents with timestamps, such as email and news, can be placed along a timeline. The timel for a set of documents returned in response to a query gives an indication of how documents relevant to that query are distributed in time. Examining the ...

Keywords: Time, ambiguity, event detection, language models, precision prediction, query classification, temporal profiles

5 LSH forest: self-tuning indexes for similarity search
Mayank Bawa, Tyson Condie, Prasanna Ganesan
May 2005 WWW '05: Proceedings of the 14th international conference on World Wide Web
Publisher: ACM
Full text available: pdf(247.91 KB) Additional Information: full citation, abstract, references, index terms

Bibliometrics: Downloads (6 Weeks): 11, Downloads (12 Months): 115, Citation Count: 6

We consider the problem of indexing high-dimensional data for answering (approximate) similar search queries. Similarity indexes prove to be important in a wide variety of settings: Web searc engines desire fast, parallel, main-memory-based indexes ...

Keywords: peer-to-peer (P2P), similarity indexes

6 Detecting near-duplicates for web crawling
Gurmeet Singh Manku, Arvind Jain, Anish Das Sarma
May 2007 WWW '07: Proceedings of the 16th international conference on World Wide Web
Publisher: ACM
Full text available: pdf(170.06 KB) Additional Information: full citation, abstract, references, index terms

Bibliometrics: Downloads (6 Weeks): 24, Downloads (12 Months): 468, Citation Count: 1

Near-duplicate web documents are abundant. Two such documents differ from each other in a v small portion that displays advertisements, for example. Such differences are irrelevant for web search. So the quality of a web crawler increases if it can ...

Keywords: fingerprint, hamming distance, near-duplicate, search, similarity, sketch, web crawl web document

7 Dynamic extraction topic descriptors and discriminators: towards automatic context-based topic search
Ana Maguitman, David Leake, Thomas Reichherzer, Filippo Menczer
November 2004 CIKM '04: Proceedings of the thirteenth ACM international conference on Informati and knowledge management
Publisher: ACM
Full text available: pdf(253.70 KB) Additional Information: full citation, abstract, references, cited by, index term

Bibliometrics: Downloads (6 Weeks): 10, Downloads (12 Months): 91, Citation Count: 1

Effective knowledge management may require going beyond initial knowledge capture, to suppo decisions about how to extend previously-captured knowledge. Electronic <i>concept maps,</i interlinked with other concept maps and multimedia resources, ...

Keywords: acquisition tools, automatic topic search, concept mapping, context, information retrieval, knowledge, knowledge management

8 Minimal document set retrieval
Wei Dai, Rohini Srihari

October 2005 CIKM '05: Proceedings of the 14th ACM international conference on Information and knowledge management
Publisher: ACM
Full text available: pdf(234.22 KB) Additional Information: full citation, abstract, references, cited by, index term

Bibliometrics: Downloads (6 Weeks): 5, Downloads (12 Months): 82, Citation Count: 3

This paper presents a novel formulation and approach to the *minimal document set retrieval* problem. Minimal Document Set Retrieval (MDSR) is a promising information retrieval task in wl each query topic is assumed to have different subtopics; ...

Keywords: document set retrieval, information retrieval

9 Cluster-based retrieval using language models
Xiaoyong Liu, W. Bruce Croft
July 2004 SIGIR '04: Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval
Publisher: ACM
Full text available: pdf(248.27 KB) Additional Information: full citation, abstract, references, cited by, index term

Bibliometrics: Downloads (6 Weeks): 24, Downloads (12 Months): 195, Citation Count: 31

Previous research on cluster-based retrieval has been inconclusive as to whether it does bring improved retrieval effectiveness over document-based retrieval. Recent developments in the language modeling approach to IR have motivated us to re-examine ...

Keywords: cluster model, cluster-based language model, cluster-based retrieval, hierarchical clustering, information retrieval, language model, query-specific clustering, smoothing, static clustering, topic model

10 Exploiting correlated keywords to improve approximate information filtering
Christian Zimmer, Christos Tryfonopoulos, Gerhard Weikum
July 2008 SIGIR '08: Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval
Publisher: ACM
Full text available: pdf(510.74 KB) Additional Information: full citation, abstract, references, index terms

Bibliometrics: Downloads (6 Weeks): 61, Downloads (12 Months): 56, Citation Count: 0

Information filtering, also referred to as publish/subscribe, complements one-time searching sin users are able to subscribe to information sources and be notified whenever new documents of interest are published. In approximate information filtering ...

Keywords: Peer-to-Peer (P2P), approximate publish/subscribe, distinct-value (DV) estimation, distributed information filtering (IF), information systems

11 Comparison of two approaches to building a vertical search tool: a case study in the nanotechnology domain
Michael Chau, Hsinchun Chen, Jialun Qin, Yilu Zhou, Yi Qin, Wai-Ki Sung, Daniel McDonald
July 2002 JCDL '02: Proceedings of the 2nd ACM/IEEE-CS joint conference on Digital libraries
Publisher: ACM
Full text available: pdf(859.29 KB) Additional Information: full citation, abstract, references, cited by, index term

Bibliometrics: Downloads (6 Weeks): 7, Downloads (12 Months): 81, Citation Count: 7

As the Web has been growing exponentially, it has become increasingly difficult to search for desired information. In recent years, many domain-specific (vertical) search tools have been developed to serve the information needs of specific fields. This ...

Keywords: indexing, information retrieval, internet searching and browsing, internet spider, no phrasing, personalization, post-retrieval analysis, self-organizing map, summarization, vertical search engine, web search engine

12 Extracting redundancy-aware top-k patterns
Dong Xin, Hong Cheng, Xifeng Yan, Jiawei Han
August 2006 KDD '06: Proceedings of the 12th ACM SIGKDD international conference on Knowledg
discovery and data mining
Publisher: ACM
Full text available: pdf(830.88 KB) Additional Information: full citation, abstract, references, index terms

Bibliometrics: Downloads (6 Weeks): 15, Downloads (12 Months): 117, Citation Count: 3

Observed in many applications, there is a potential need of extracting a small set of frequent patterns having not only high significance but also low redundancy. The significance is usually defined by the context of applications. Previous studies have ...

Keywords: pattern extraction, redundancy, significance

13 Computer-based plagiarism detection methods and tools: an overview
Romans Lukashenko, Vita Graudina, Janis Grundspenkis
June 2007 CompSysTech '07: Proceedings of the 2007 international conference on Computer
systems and technologies
Publisher: ACM
Full text available: pdf(94.34 KB) Additional Information: full citation, abstract, references

Bibliometrics: Downloads (6 Weeks): 12, Downloads (12 Months): 222, Citation Count: 0

The paper is dedicated to plagiarism problem. The ways how to reduce plagiarism: both: plagiar prevention and plagiarism detection are discussed. Widely used plagiarism detection methods ar described. The most known plagiarism detection tools are ...

Keywords: plagiarism, plagiarism detection, plagiarism prevention, similarity measures

14 Finding similar experts
Krisztian Balog, Maarten de Rijke
July 2007 SIGIR '07: Proceedings of the 30th annual international ACM SIGIR conference on
Research and development in information retrieval
Publisher: ACM
Full text available: pdf(270.49 KB) Additional Information: full citation, abstract, references, index terms

Bibliometrics: Downloads (6 Weeks): 13, Downloads (12 Months): 172, Citation Count: 0

The task of finding people who are experts on a topic has recently received increased attention. introduce a different expert finding task for which a small number of example experts is given (instead of a natural language query), and the system's ...

Keywords: expert finding, expert representation, similar experts

15 Detection of Duplicate Defect Reports Using Natural Language Processing
Per Runeson, Magnus Alexandersson, Oskar Nyholm
May 2007 ICSE '07: Proceedings of the 29th international conference on Software Engineering
Publisher: IEEE Computer Society
Full text available: pdf(268.53 KB) Additional Information: full citation, abstract, references, index terms

Bibliometrics: Downloads (6 Weeks): 2, Downloads (12 Months): 225, Citation Count: 2

Defect reports are generated from various testing and development activities in software engineering. Sometimes two reports are submitted that describe the same problem, leading to duplicate reports. These reports are mostly written in structured natural ...

16 THESUS: Organizing Web document collections based on link semantics

Maria Halkidi, Benjamin Nguyen, Iraklis Varlamis, Michalis Vazirgiannis
November 2003 The VLDB Journal — The International Journal on Very Large Data Bases, Volume 12 Issue 4
Publisher: Springer-Verlag New York, Inc.
Full text available: pdf(262.85 KB) Additional Information: full citation, abstract, references, cited by, index term

Bibliometrics: Downloads (6 Weeks): 9, Downloads (12 Months): 90, Citation Count: 5

The requirements for effective search and management of the WWW are stronger than ever. Currently Web documents are classified based on their content not taking into account the fact t these documents are connected to each other by links. We claim ...

Keywords: Document clustering, Link analysis, Link management, Semantics, Similarity measu World Wide Web

17 Metric space similarity joins
Edwin H. Jacox, Hanan Samet
June 2008 ACM Transactions on Database Systems (TODS), Volume 33 Issue 2
Publisher: ACM
Full text available: pdf(1.10 MB) Additional Information: full citation, abstract, references, index terms

Bibliometrics: Downloads (6 Weeks): 52, Downloads (12 Months): 118, Citation Count: 0

Similarity join algorithms find pairs of objects that lie within a certain distance &epsi; of each other. Algorithms that are adapted from spatial join techniques are designed primarily for data i vector space and often employ some form of a multidimensional ...

Keywords: Similarity join, distance-based indexing, external memory algorithms, nearest neighbor queries, range queries, ranking

18 Efficient similarity joins for near duplicate detection
Chuan Xiao, Wei Wang, Xuemin Lin, Jeffrey Xu Yu
April 2008 WWW '08: Proceeding of the 17th international conference on World Wide Web
Publisher: ACM
Full text available: pdf(327.62 KB) Additional Information: full citation, abstract, references, index terms

Bibliometrics: Downloads (6 Weeks): 33, Downloads (12 Months): 118, Citation Count: 0

With the increasing amount of data and the need to integrate data from multiple data sources, a challenging issue is to find near duplicate records efficiently. In this paper, we focus on efficient algorithms to find pairs of records such that their ...

Keywords: near duplicate detection, similarity join

19 Communications of the ACM: Volume 51 Issue 1
January 2008 issue Volume 51 Issue 1
Publisher: ACM
Full text available: pdf(5.97 MB) digital edition Additional Information: full citation, index terms

Bibliometrics: Downloads (6 Weeks): 553, Downloads (12 Months): 3458, Citation Count: 0

20 A survey of Web metrics
Devanshu Dhyani, Wee Keong Ng, Sourav S. Bhowmick
December 2002 ACM Computing Surveys (CSUR), Volume 34 Issue 4
Publisher: ACM
Full text available: pdf(289.28 KB) Additional Information: full citation, abstract, references, cited by, index term

Bibliometrics: Downloads (6 Weeks): 70, Downloads (12 Months): 707, Citation Count: 18

The unabated growth and increasing significance of the World Wide Web has resulted in a flurry research activity to improve its capacity for serving information more effectively. But at the hea of these efforts lie implicit assumptions about "quality" ...

Keywords: Information theoretic, PageRank, Web graph, Web metrics, Web page similarity, quality metrics

Results 1 - 20 of 52 Result page: 1 2 3 next >>